AIP | The Journal of Chemical Physics

Using the ONIOM hybrid method to apply equation of motion CCSD to larger systems: Benchmarking and comparison with time-dependent density functional theory, configuration interaction singles, and time-dependent Hartree–Fock

Marco Caricato, Thom Vreven, Gary W. Trucks, Michael J. Frisch, and Kenneth B. Wiberg

Citation: The Journal of Chemical Physics **131**, 134105 (2009); doi: 10.1063/1.3236938
View online: http://dx.doi.org/10.1063/1.3236938
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/131/13?ver=pdfcov
Published by the AIP Publishing

Using the ONIOM hybrid method to apply equation of motion CCSD to larger systems: Benchmarking and comparison with time-dependent density functional theory, configuration interaction singles, and time-dependent Hartree–Fock

Marco Caricato,[1,2,a)] Thom Vreven,[2] Gary W. Trucks,[2] Michael J. Frisch,[2] and Kenneth B. Wiberg[1]
[1]*Department of Chemistry, Yale University, 225 Prospect St., New Haven, Connecticut 06511, USA*
[2]*Gaussian, Inc., 340 Quinnipiac St., Bldg. 40, Wallingford, Connecticut 06492, USA*

(Received 14 July 2009; accepted 3 September 2009; published online 5 October 2009)

Equation of motion coupled-cluster singles and doubles (EOM-CCSD) is one of the most accurate computational methods for the description of one-electron vertical transitions. However, its $\mathcal{O}(N^6)$ scaling, where N is the number of basis functions, often makes the study of molecules larger than 10–15 heavy atoms prohibitive. In this work we investigate how accurately less expensive methods can approximate the EOM-CCSD results. We focus on our own N-layer integrated molecular orbital molecular mechanics (ONIOM) hybrid scheme, where the system is partitioned into regions which are treated with different levels of theory. For our set of benchmark calculations, the comparison of conventional configuration interaction singles (CIS), time-dependent Hartree–Fock (TDHF), and time-dependent density functional theory (TDDFT) methods and ONIOM (with different low level methods) showed that the best accuracy-computational time combination is obtained with ONIOM(EOM:TDDFT), which has a rms of the error with respect to the conventional EOM-CCSD of 0.06 eV, compared with 0.47 eV of the conventional TDDFT. © *2009 American Institute of Physics*. [doi:10.1063/1.3236938]

I. INTRODUCTION

The single reference coupled cluster (CC) theory, in its equation of motion (EOM-CC)[1–7] or equivalent linear response (LR-CC)[7–9] formalism, is one of the most powerful tools to calculate electronic transition energies in quantum chemistry. The most widely used formulation of the theory includes singles and doubles excitation operators only, in both the ground state and excited state expansions, EOM-CCSD.[3,4,9] With this choice of the truncation the method scales as $\mathcal{O}(N^6)$, where N is the number of basis functions. This means that it is still computationally feasible while it carries enough flexibility in the wave function to describe most of the one-electron excitations. Another remarkable characteristic of the method is that it is a *black box*, in the sense that there is no ambiguity in the definition of the wave function once one specifies geometry and basis set. Moreover, in CC theory, the quality of the results can be systematically improved, contrary to the density functional theory (DFT), by increasing the cluster expansion.

Despite it being one of the most affordable among the possible and meaningful truncations of the cluster expansion, the EOM-CCSD computational cost, along with large memory and disk requirements, prevents its use on many systems of chemical, if not biological, interest. Calculations on systems with more than 10–15 heavy atoms are prohibitive without supercomputer facilities and are not considered routine. The two traditional approaches to deal with intractable calculations are to either reduce the size of the system but still considering it at the high level of theory or reduce the level of theory but still considering the entire system. Neither is very satisfactory and often a too drastic approximation. In this work we investigate an alternative to the traditional approaches, which is to apply a hybrid method: Our own N-layer integrated molecular orbital molecular mechanics (ONIOM).[10–18] Hybrid methods divide the system into regions and treat each with a different level of theory. By only using expensive methods for the part of the system where it is necessary, hybrid methods allow the accurate study of the chemistry of very large systems. Most of those methods combine a quantum mechanical level with a molecular mechanical level, QM/MM.[19–21] Unlike those methods, that are formulated as summation, ONIOM is formulated as extrapolation. This allows the possibility of combining more than two computational levels, as well as the integration of two different QM levels, QM/QM. Also transition energies can be easily defined within the ONIOM scheme.[22]

The first important assumption in a hybrid method is that it must be possible to identify a core region, or *model* system, which is treated at the high level of theory. This can be, for example, a small chromophore in a protein or in solution (two layer scheme). A three layer scheme could include the chromophore in the first layer, some strongly interacting groups of the protein or the first solvation shell in the second layer, and the rest in a third layer. Another example could be a large chromophore that itself is too large to be treated entirely at EOM-CCSD level. Some of us showed that as

a)Electronic mail: marco.caricato@yale.edu.

long as there is a core part, ONIOM can even partition the chromophore into different regions.[18,22]

Whereas the behavior of ONIOM for ground state energies has been investigated in numerous works, there are only few reports in the literature on the calculation of transition energies,[22–34] of which several used QM/QM combinations. The goal of this work is to investigate further the effectiveness of ONIOM(QM:QM) for the description of excited states. In particular, we attempt to reproduce the EOM-CCSD energy calculated on selected test cases. With EOM-CCSD as high level method, we investigated various low level methods, basis sets and partitions in order to study a representative sample of possibilities and to recommend ONIOM schemes for future production calculations. The reliability of the ONIOM results is primarily determined by two factors: The high level model needs to capture most of the transition process, while the low level of theory should describe the remaining contributions as well as the high level does. Although a good balance of these two points and the computational cost requires a calibration step that depends on the particular system under study, the current work will provide some general guidelines. In this report we focus only on the energy since this is the fundamental quantity that ONIOM must be able to reproduce in order to be a valuable method in production calculations. The best strategy for the definition of other properties in the ONIOM framework, for instance, oscillator strengths or excited state dipole moments, has not been properly investigated yet (integration of the densities or extrapolation of the property, for example) and will be material for future research.

This paper is organized as follows. Section II reviews the theory of the ONIOM method and reports the computational details. Section III collects the results for the test cases we examined and discusses the aspects that influence the ONIOM final results. A summary of the discussions and general conclusions are reported in Sec. IV.

II. THEORY AND METHODS

The ONIOM energy for a two layer system is written as an extrapolation,

$$E^{\mathrm{ONIOM}} = E^{\mathrm{high}}_{\mathrm{model}} + E^{\mathrm{low}}_{\mathrm{real}} - E^{\mathrm{low}}_{\mathrm{model}}, \tag{1}$$

where *real* and *model* refer to the full system and to the core region, respectively. The coupling between the layers is included at the low level of theory. As evident from Eq. (1), the ONIOM single-point energy can be calculated by combining the results of three separate energy calculations.

When there is covalent bonding between the two regions, open valencies result in the model system, which are saturated with (hydrogen) link atoms. The bond length of the link atom is obtained by scaling the original bond length of the atom that is substituted with a factor that depends on the typical bond distance of the atoms involved. The exact value of the scale factor has a very little effect on the absolute ONIOM energy and even less on geometry optimizations.[35] There are simple rules that should be followed in the definition of the model system. The cuts should be made on single, nonpolar, and nonconstrained bonds. Also, an atom in the low level layer should not be bonded to two or more atoms in the high level layer (resulting in link atoms too close to each other). Respecting these partitioning rules helps reduce the errors in the extrapolation scheme and obtain meaningful results.

The transition energy in the ONIOM scheme can be expressed as the difference of the ONIOM energies of the two states,

$$\begin{aligned}
\Delta E^{\mathrm{ONIOM}} &= E^{*,\mathrm{ONIOM}} - E^{\mathrm{ONIOM}} \\
&= (E^{*,\mathrm{high}}_{\mathrm{model}} + E^{*,\mathrm{low}}_{\mathrm{real}} - E^{*,\mathrm{low}}_{\mathrm{model}}) \\
&\quad - (E^{\mathrm{high}}_{\mathrm{model}} + E^{\mathrm{low}}_{\mathrm{real}} - E^{\mathrm{low}}_{\mathrm{model}}) \\
&= (E^{*,\mathrm{high}}_{\mathrm{model}} - E^{\mathrm{high}}_{\mathrm{model}}) + (E^{*,\mathrm{low}}_{\mathrm{real}} - E^{\mathrm{low}}_{\mathrm{real}}) \\
&\quad - (E^{*,\mathrm{low}}_{\mathrm{model}} - E^{\mathrm{low}}_{\mathrm{model}}) \\
&= \Delta E^{\mathrm{high}}_{\mathrm{model}} + \Delta E^{\mathrm{low}}_{\mathrm{real}} - \Delta E^{\mathrm{low}}_{\mathrm{model}}.
\end{aligned} \tag{2}$$

Equation (2) shows that also the ONIOM transition energy is the combination of three separate transition energy calculations, similarly to Eq. (1).

In contrast to many ground state phenomena, excitations are often more delocalized. Molecular mechanics methods cannot describe excitations. Therefore, in QM/MM calculations, including ONIOM(QM:MM), the excitation is forced to be localized in the QM region. $\Delta E^{\mathrm{low}}_{\mathrm{real}}$ and $\Delta E^{\mathrm{low}}_{\mathrm{model}}$ in Eq. (2) are then zero. However, when the low level of theory is a QM method that can describe electronic excitations, these two terms can be nonzero. This allows the excitation to extend into the low level region in ONIOM(QM:QM) calculations. Another complication with respect to ground state calculations is that the ordering of the states is not necessarily the same for the three terms in Eq. (2). Thus one must ensure to combine states of the same character in order to have an appropriate description of the excitation. In other words, the three calculations in Eq. (2) must qualitatively describe the same transition.

The ONIOM success depends on the accuracy of the results and on the computational effort in comparison to the target calculation (defined as the high level of theory on the entire system). Among the various choices that influence the final ONIOM output and that we analyze here are the low level method and basis set, the size of the model system, and the link atoms effect. We also compare the target with the results of the $\Delta E^{\mathrm{high}}_{\mathrm{model}}$ and the $\Delta E^{\mathrm{low}}_{\mathrm{real}}$ terms alone. This verifies if and how ONIOM improves the traditional approximations by introducing the substituent effects in the first case and a higher level of theory in the chromophore region in the second case.

We do not present detailed CPU timings as they heavily depend on the machine setup, and we limit the comparison of the different terms of the ONIOM calculation with the target. Our goal is to give a feeling of the difference between the target and the ONIOM CPU times. The bottleneck in the ONIOM calculations presented here is usually the $\Delta E^{\mathrm{high}}_{\mathrm{model}}$ term, but sometimes the $\Delta E^{\mathrm{low}}_{\mathrm{real}}$ can also be significant depending on the relative size of the real and the model systems and on the relative cost of the high and low level. Whereas the limiting factor in most QM calculations is the

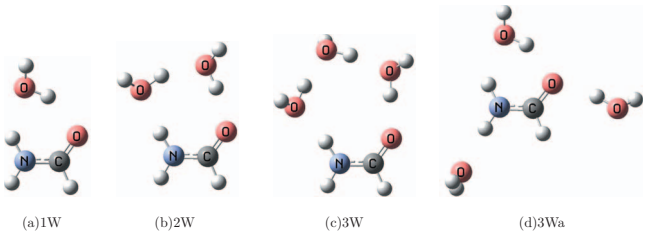


FIG. 1. Structures of the four water clusters.

computational time, EOM-CCSD calculations on large molecules can be limited by memory and disk requirements as well. We do not investigate this aspect in detail but want to note that the storage requirements in an ONIOM calculation with EOM-CCSD can be reduced dramatically.

The definition of the error used throughout the paper is

$$\text{Error} = \Delta E^{\text{target}} - X; \quad X = \Delta E^{\text{ONIOM}}, \Delta E^{\text{low}}_{\text{real}}, \Delta E^{\text{high}}_{\text{model}}. \quad (3)$$

The target level of theory in this work is EOM-CCSD on the full system with the largest basis set (6-311+G(d,p)) we consider. The ONIOM potential is always supposed to be an approximation to the high level potential. Since we are attempting to reproduce EOM-CCSD/6-311+G(d,p) calculations, we restrict ourselves to this level of theory for the high level in ONIOM. If ONIOM calculations with methods other than this one in the high level were to reproduce the target better, this would almost certainly be the result of cancellation of errors and not provide the insight that we are trying to obtain with this work.

We only consider vertical transitions from the ground state geometries. These are optimized at the conventional B3LYP/6-311+G(d,p) level on the entire system, and they are used for all the transition energy calculations, both ONIOM and conventional.

The low level methods considered are EOM-CCSD (with basis sets smaller than the target), time-dependent DFT (TDDFT) with the B3LYP hybrid functional,[36–38] configuration interaction singles (CIS), and time-dependent Hartree–Fock (TDHF). Functionals other than B3LYP may be more appropriate for specific cases, but finding the best functional for each system is beyond the scope of this work.

Three basis sets are investigated for each low level method and denoted *large*, *medium*, and *small*. 6-311+G(d,p) is the large basis set (L), 6-31+G(d) is the medium (M), and 6-31G(d) is the small (S) basis set. 6-31+G(d) is chosen as medium because it requires a smaller computational effort with respect to large but retains diffuse functions, which may be important for a qualitatively correct description of electronic transitions. All the calculations are performed with a development version of the Gaussian suite of programs[39] in which the EOM-CCSD method has been recently implemented.

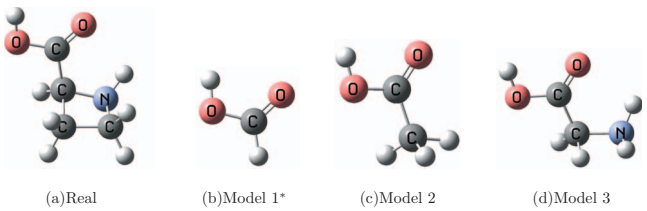


FIG. 2. Structures of *S*-azetidine-2-carboxylic acid and of the three model systems. The (*) indicates the preferred model system.

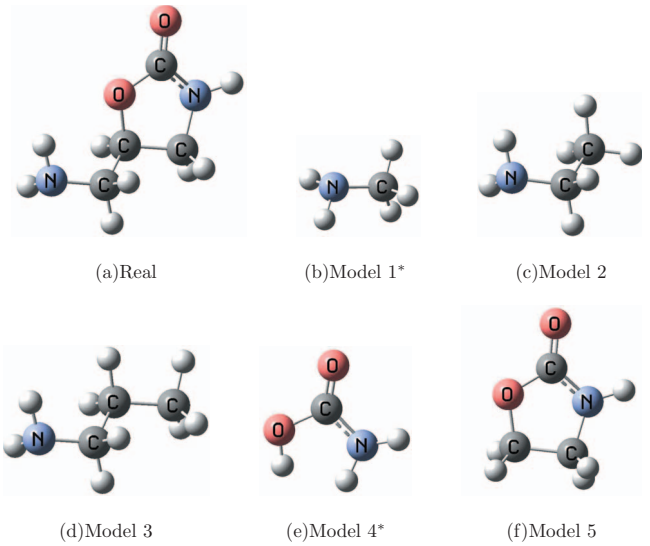


FIG. 3. Structures of *R*-5-aminomethyl-2-oxazolidinone and of the five model systems. Models 1, 2, and 3 are used to describe the first transition; models 4 and 5 describe the second transition. The (*) indicates the preferred model system.

III. RESULTS AND DISCUSSION

A. Test systems

We use a series of five test systems to assess the methods. The choice of the tests is made in order to deal with a variety of possible causes of errors introduced by a hybrid method such as ONIOM. Furthermore, the systems need to be small enough to be treated entirely at the high level of theory but sufficiently large to allow for a reasonable partitioning into high and low level regions. The limited size of the molecules often forces us to "cut" at places that are not usually recommended in ONIOM calculations, such as

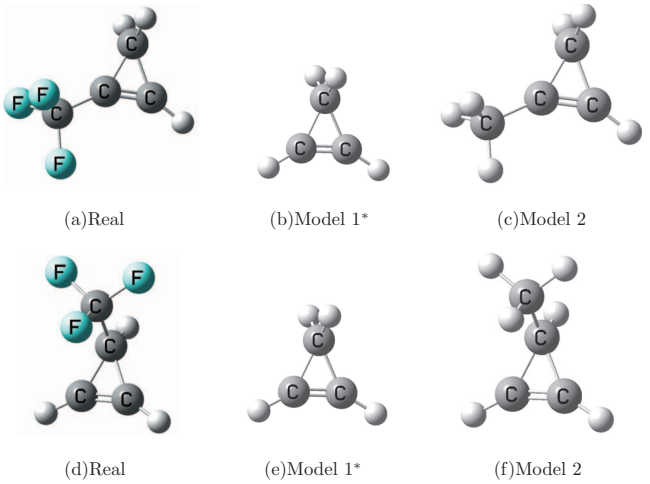
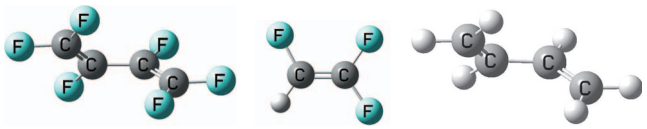


FIG. 4. Structures of 1-trifluoro-methyl-cyclopropene and of 3-trifluoro-methyl-cyclopropene and their model systems. The (*) indicates the preferred model system.



(a)Real (b)Model 1 (c)Model 2

FIG. 5. Structures of hexafluoro-1,3-butadiene and of the two model systems.

across polar bonds or through cyclic structures (see Sec. II). Since these partitions are more drastic than what is typical in production calculations, they provide stringent tests of the limits of this methodology. The first one or two excited states are considered for each molecule, which are valence in character. These are in fact the states of interest in most studies that involve large chromophores, which are the usual target of ONIOM production calculations. However, excited states of different nature, for instance Rydberg or charge transfer, will be subject of future research.

The five systems we study are various water clusters around a formamide molecule, *S*-azetidine-2-carboxylic acid, *R*-5-aminomethyl-2-oxazolidinone, two trifluoro-methyl-cyclopropenes, and hexafluoro-1,3-butadiene. The structures of these systems and of the related models are reported in Figs. 1–5. The CPU time speed-ups for the most expensive terms of the ONIOM energy are reported in Table I for three test cases. The speed-up is relative to the target calculation considered as unity. The errors, as defined in Eq. (3), are reported in Tables II–VIII.[40] Relative errors (%) are often reported in parentheses in the text. A graphical representation of the errors for the formamide-water clusters, *S*-azetidine-2-carboxylic acid, and *R*-5-aminomethyl-2-oxazolidinone is reported in Figs. 6–10. The root mean square (rms) of the errors and the maximum absolute error (MAE) for the preferred ONIOM models and the stand alone calculations, $\Delta E^{\text{low}}_{\text{real}}$ and $\Delta E^{\text{high}}_{\text{model}}$, are reported in Table IX [all the errors refer to the target values, Eq. (3)]. All the test molecules are included except for the hexafluoro-1,3-butadiene. This molecule cannot be effectively treated by the ONIOM hybrid method, as discussed in Sec. III B 5, so its inclusion in the overall errors would be misleading.

TABLE I. CPU time speed-ups for the most expensive terms of the ONIOM energy for three of the test cases. The speed-up is relative to the target calculations.

	Formamide-water clusters		
	1W	2W	3W[a]
$\Delta E^{\text{high}}_{\text{model}}$	5	17	45
$\Delta E^{\text{low}}_{\text{real}}$ [b]	2.4	3	3
$\Delta E^{\text{high}}_{\text{model-1}}$	$\Delta E^{\text{high}}_{\text{model-2}}$	$\Delta E^{\text{high}}_{\text{model-3}}$	$\Delta E^{\text{low}}_{\text{real}}$ [b]
	S-azetidine-2-carboxylic acid		
29	7	3	2.5
	R-5-aminomethyl-2-oxazolidinone, first transition		
180	36	12	8

[a]3W and 3Wa have comparable CPU timing; thus only one is reported.
[b]EOM/M level.

B. Specific systems

1. Formamide-water clusters

Four clusters are considered, which include one, two, and three water molecules, the latter in two different conformations. The clusters, shown in Fig. 1, are named 1W, 2W, 3W, and 3Wa, respectively. We consider the first excitation ($n \rightarrow \pi^*$). For these systems the obvious ONIOM partition is to consider the formamide as the model system and the water molecules in the low level region. This partition does not require the use of link atoms as no covalent bond is cut. We note that even if the model system is identical for the four clusters, its geometry differs in the four cases as it is determined with the different conformations of water molecules.

Starting with the low cost methods, the best conventional results are from TDDFT/M and S, with absolute errors below 0.1 eV (0%–2%), except for 3Wa, where the error is 0.12 eV, as shown in Fig. 6 and Table II. However, this good performance seems fortuitous as TDDFT/L has an error that lies around 0.15–0.21 eV (3%). Indeed, the transition energy increases with decreasing size of the basis set, which then cancels with the underestimation of ΔE by TDDFT. The smallest $\Delta E^{\text{low}}_{\text{real}}$ errors are thus obtained with TDDFT/L if we do not take into account the "lucky case" of TDDFT/M and S. The $\Delta E^{\text{high}}_{\text{model}}$ error increases with the size of the cluster (5%–10%), as expected, although cluster 2W and 3W are very close (6%) because the third water molecule in 3W does not interact directly with the high level region.

The ONIOM error, with all the low level methods, is smaller than the conventional calculations ($\Delta E^{\text{low}}_{\text{real}}$ and $\Delta E^{\text{high}}_{\text{model}}$), with the only exception of TDDFT/S for 2W, 3W, and 3Wa (see Table II). ONIOM TDDFT/S provides the best results, but this is likely due to error cancellation, as discussed above. More reliable are the results obtained using TDDFT/L and M in the low level, with errors less than 0.1 eV (1%). Both CIS and TDHF as low levels are worse than TDDFT (errors of around 1%–3%).

The conventional EOM-CCSD provides small $\Delta E^{\text{low}}_{\text{real}}$ errors (2%–4%), but the ONIOM errors with this method in the low level are even smaller (0%–1%), especially with the M basis set (see Table II). In addition, the error with EOM-CCSD is less sensitive to the size of the clusters. ONIOM with EOM-CCSD/M yields the closest result to the target.

The CPU timing is reported in Table I. The computational time for $\Delta E^{\text{high}}_{\text{model}}$ is roughly the same for the four clusters as the models differ only in the geometry. This is the slowest step for the ONIOM calculations with the low cost methods. The ONIOM calculations are considerably faster than the target, and the difference increases with the size of the cluster. On the other hand, $\Delta E^{\text{low}}_{\text{real}}$ becomes the bottleneck with EOM-CCSD as low level.

2. S-azetidine-2-carboxylic acid

S-azetidine-2-carboxylic acid is a non-natural amino acid that has chemical and biochemical applications[41–46] [see Fig. 2(a)]. The first transition in this system mainly involves the carboxylic group. We define three models of increasing size,

TABLE II. Target values, $\Delta E_{\text{real}}^{\text{low}}$, and ONIOM errors (eV) for the water clusters in Fig. 1. The numbers in parentheses are the $\Delta E_{\text{model}}^{\text{high}}$ errors.

	$\Delta E_{\text{real}}^{\text{low}}$				ONIOM			
	1W	2W	3W	3Wa	1W	2W	3W	3Wa
Target	5.91	6.00	6.00	6.16	(0.27)	(0.38)	(0.38)	(0.61)
EOM/M	−0.09	−0.09	−0.09	−0.09	0.00	0.00	0.01	0.00
EOM/S	−0.20	−0.20	−0.20	−0.22	−0.03	−0.03	−0.02	−0.04
TDDFT/L	0.15	0.18	0.19	0.21	0.03	0.06	0.06	0.09
TDDFT/M	0.06	0.09	0.09	0.12	0.04	0.06	0.07	0.10
TDDFT/S	−0.04	−0.01	0.00	0.04	0.01	0.03	0.04	0.08
TDHF/L	−0.75	−0.79	−0.80	−0.87	−0.09	−0.13	−0.14	−0.18
TDHF/M	−0.82	−0.87	−0.88	−0.95	−0.09	−0.12	−0.13	−0.18
TDHF/S	−0.86	−0.91	−0.92	−0.99	−0.10	−0.14	−0.15	−0.19
CIS/L	−0.90	−0.94	−0.95	−1.01	−0.08	−0.11	−0.12	−0.15
CIS/M	−0.97	−1.01	−1.02	−1.08	−0.08	−0.10	−0.11	−0.15
CIS/S	−1.02	−1.06	−1.07	−1.13	−0.09	−0.12	−0.13	−0.16

shown in Figs. 2(b)–2(d). Here the models are built by cutting covalent bonds and, for models 2 and 3, involve polar bonds and cyclic structures (see Sec. II).

All the conventional HF based methods overestimate the target, including EOM-CCSD (see Fig. 8 and Table III). Both CIS and TDHF are very far from the target, with TDHF (12%–13%) being slightly better than CIS (15%). Conventional TDDFT on the other hand underestimates the transition energy (0%–2%). The $\Delta E_{\text{model}}^{\text{high}}$ error decreases with increasing size of the model (3→0%). For model 3, the *high* term $\Delta E_{\text{model–3}}^{\text{high}}$ is very close to the target, but this is also the computationally most expensive model.

For ONIOM with model 1 (ONIOM-1) with the low cost methods, the best results are obtained with TDDFT as the low level, as shown in Table III and Fig. 8. For ONIOM-2 and -3, the improvement over the $\Delta E_{\text{model}}^{\text{high}}$ calculation is expected to be relatively small as the $\Delta E_{\text{model}}^{\text{high}}$ error is already small, 0.05 and −0.02 eV, respectively. With ONIOM-2, both CIS and TDHF with the L and M basis sets perform very well. On the other hand, the TDDFT performance using this model is slightly worse. All the ONIOM-3 errors are larger than $\Delta E_{\text{model–3}}^{\text{high}}$, and yet, overall, they remain small.

The conventional EOM-CCSD with the medium basis set is quite close to the target with an error of −0.07 eV (1%), whereas the S basis set yields an error of −0.14 eV (2%). However, EOM-CCSD as low level in ONIOM, with both M and S basis sets, provides extremely good results, allowing for an improvement even over $\Delta E_{\text{model–2}}^{\text{high}}$ and $\Delta E_{\text{model–3}}^{\text{high}}$.

TABLE III. Target value, $\Delta E_{\text{real}}^{\text{low}}$, and ONIOM errors (eV) for *S*-azetidine-2-carboxylic acid in Fig. 2. The numbers in parentheses are the $\Delta E_{\text{model}}^{\text{high}}$ errors.

	$\Delta E_{\text{real}}^{\text{low}}$	ONIOM-1	ONIOM-2	ONIOM-3
Target	5.93	(0.16)	(0.05)	(−0.02)
EOM/M	−0.07	0.02	0.02	0.00
EOM/S	−0.14	0.02	−0.02	0.01
TDDFT/L	0.10	−0.02	−0.07	−0.05
TDDFT/M	0.06	0.03	−0.05	−0.05
TDDFT/S	0.03	0.04	−0.09	−0.04
TDHF/L	−0.70	−0.15	−0.03	−0.02
TDHF/M	−0.75	−0.11	−0.02	−0.01
TDHF/S	−0.74	−0.09	−0.05	−0.02
CIS/L	−0.87	−0.13	−0.02	−0.02
CIS/M	−0.91	−0.10	−0.02	−0.01
CIS/S	−0.91	−0.08	−0.04	−0.02

From the timing in Table I it is evident that model 1 represents a drastic speed-up with respect to the target calculation (~29 times). Thus an ONIOM calculation that includes this model with any of the low cost methods in the low level implies a dramatic reduction in the computational cost. With EOM-CCSD/M as low level, the $\Delta E_{\text{real}}^{\text{low}}$ term is the bottleneck of the calculation, although the computational time is still significantly reduced. A balanced ratio of CPU time saving and accuracy for this system is obtained by using model 1 and TDDFT/L at low level.

3. R-5-aminomethyl-2-oxazolidinone

R-5-aminomethyl-2-oxazolidinone [Fig. 3(a)] belongs to a class of substances with antibacterial and antimicrobial activities.[47–49] The first two electronic transitions are localized onto two different groups, and we show how they can be studied by selecting different model systems. Here we also consider the difference in the transition energies between the two excited states, $\Delta E_{1\to 2}=\Delta E_2-\Delta E_1$.

The first transition is mainly localized around the NH_2 group, and we define three model systems, increasing in size, shown in Figs. 3(b)–3(d). As in Sec. III B 2, model 2 and 3 require "cuts" that are usually not recommended.

Conventional TDDFT/L and M underestimate the target energy by ~0.5 eV (8%), whereas CIS and TDHF with the same basis sets, which are very close to each other, overestimate it by about 1.16–1.19 eV (18%–19%). The $\Delta E_{\text{model}}^{\text{high}}$ errors are quite large (4%–6%), as shown in Table IV. The error for model 2 is smaller than for model 3, despite the latter including a larger part of the entire system. However, the unusual partitioning in those models can imply less predictable effects. Table IV and Fig. 9 show that all the ONIOM errors are smaller than the associated stand alone calculations. The best ONIOM-1 results are obtained with

TABLE IV. Target values, $\Delta E^{\text{low}}_{\text{real}}$, and ONIOM errors (eV) for the two transitions of *R*-5-aminomethyl-2-oxazolidinone in Fig. 3. The numbers in parentheses are the $\Delta E^{\text{high}}_{\text{model}}$ errors.

	First transition				Second transition		
	$\Delta E^{\text{low}}_{\text{real}}$	ONIOM-1	ONIOM-2	ONIOM-3	$\Delta E^{\text{low}}_{\text{real}}$	ONIOM-4	ONIOM-5
Target	6.32	(0.40)	(0.25)	(0.37)	6.71	(−0.05)	(0.03)
EOM/M	0.06	0.01	0.05	0.00	0.04	0.06	0.01
EOM/S	−1.78	0.03	−0.02	0.04	−1.74	−0.71	0.05
TDDFT/L	0.49	0.04	0.04	0.05	0.37	−0.03	−0.02
TDDFT/M	0.53	0.06	0.09	0.06	0.38	0.01	0.00
TDDFT/S	−0.84	0.29	0.18	0.23	−1.12	−0.94	−0.03
TDHF/L	−1.16	−0.11	−0.04	−0.03	−1.19	0.00	0.00
TDHF/M	−1.17	−0.09	0.00	−0.04	−1.22	0.01	0.01
TDHF/S	−2.72	−0.02	−0.04	0.03	−2.80	−0.71	0.03
CIS/L	−1.18	−0.12	−0.04	−0.03	−1.21	0.02	0.00
CIS/M	−1.19	−0.09	0.00	−0.03	−1.24	0.02	0.01
CIS/S	−2.77	−0.03	−0.04	0.03	−2.91	−0.76	0.01

TDDFT/L, with an error of 0.04 eV (1%). For ONIOM-2 and -3, all the methods with the L basis set reduce the error in $\Delta E^{\text{high}}_{\text{model}}$ by a factor of eight to ten; when the smaller basis sets are used, there is always improvement over $\Delta E^{\text{high}}_{\text{model}}$, but the remaining errors are less constant (0%–4%).

EOM-CCSD/M presents a small $\Delta E^{\text{low}}_{\text{real}}$ error (1%), which slightly underestimates the target (0.06 eV). However, the use of the ONIOM approach manages to reduce this already small error. Very good results are obtained also with the S basis set.

All the partitions represent a considerable improvement with respect to the target as far as the CPU time is concerned (see Table I). However, the calculation of $\Delta E^{\text{high}}_{\text{model}-1}$ is drastically faster than all the others model systems. Thus, when a method such as TDDFT/L is used in the low level, ONIOM-1 yields a very accurate estimate of the target calculation at a very small computational cost. When EOM-CCSD/M is used as low level, the slowest step is the $\Delta E^{\text{low}}_{\text{real}}$ term, independently on the model. Still, the reduction of the computational time is significant.

The second electronic transition involves the CO_2N group. We define two models, shown in Figs. 3(e) and 3(f). Model 4 includes only the group responsible for the excitation; model 5 keeps the complete ring, eliminating only two heavy atoms. All the conventional methods present very large deviations from the target (17%–43%), except EOM-CCSD/M, TDDFT/L, and M (0.04, 0.37, and 0.38 eV, 1% and 6%, respectively, Table IV and Fig. 10). The $\Delta E^{\text{high}}_{\text{model}}$ errors are very small (1%), indicating that both models contain enough information on the transition, i.e., the substituent effect is small. With ONIOM the errors are always small, except for model 4 and the S basis set (see Fig. 10). To conclude, this example shows, on one hand, that a proper basis set is important at low level and, on the other hand, that even if the $\Delta E^{\text{high}}_{\text{model}}$ calculation already provides a good estimate of the target transition energy, the ONIOM extrapolation keeps the error small or even reduces it if the conventional rules of partitioning are followed.

For the energy difference between the two excited states, $\Delta E_{1\to 2}$, we compare all the ONIOM combinations for the two transitions, as well as the stand alone calculations. This is effectively an ONIOM calculation with two high level

TABLE V. Target value and errors for the transition energy difference $\Delta E_{1\to 2}$ (eV) between the second and the first excited states of *R*-5-aminomethyl-2-oxazolidinone calculated at $\Delta E^{\text{low}}_{\text{real}}$ level and with the various combinations of ONIOM models for the two transitions. The numbers in parentheses are the $(\Delta E^{\text{high}}_{\text{model}})_{1\to 2}$ errors. The labels refer to the models in Fig. 3. For example, 4-1 denotes that model 4 is used for the second excitation and model 1 is used for the first excitation.

	Real	4-1	4-2	4-3	5-1	5-2	5-3
Target	0.39	(−0.44)	(−0.29)	(−0.42)	(−0.36)	(−0.21)	(−0.34)
EOM/M	−0.02	0.05	0.01	0.06	0.00	−0.04	0.01
EOM/S	0.04	−0.75	−0.69	−0.76	0.02	0.07	0.01
TDDFT/L	−0.12	−0.07	−0.07	−0.08	−0.05	−0.06	−0.07
TDDFT/M	−0.15	−0.05	−0.08	−0.04	−0.07	−0.10	−0.06
TDDFT/S	−0.28	−1.22	−1.11	−1.17	−0.31	−0.20	−0.26
TDHF/L	−0.03	0.11	0.04	0.03	0.12	0.05	0.04
TDHF/M	−0.05	0.10	0.02	0.05	0.10	0.01	0.05
TDHF/S	−0.08	−0.68	−0.67	−0.74	0.05	0.07	0.00
CIS/L	−0.03	0.13	0.05	0.05	0.12	0.04	0.04
CIS/M	−0.05	0.11	0.03	0.06	0.10	0.01	0.04
CIS/S	−0.14	−0.73	−0.72	−0.79	0.04	0.06	−0.01

TABLE VI. Target values, $\Delta E_{\text{real}}^{\text{low}}$, and ONIOM $\sigma \rightarrow \pi^*$ transition energy errors (eV) for the two isomers of the substituted cyclopropene in Fig. 4. The numbers in parentheses are the $\Delta E_{\text{model}}^{\text{high}}$ errors.

	1-trifluoro-methyl-cyclopropenes			3-trifluoro-methyl-cyclopropenes		
	$\Delta E_{\text{real}}^{\text{low}}$	ONIOM-1	ONIOM-2	$\Delta E_{\text{real}}^{\text{low}}$	ONIOM-1	ONIOM-2
Target	6.88	(−0.03)	(−0.21)	7.17	(0.32)	(0.68)
EOM/M	−0.06	0.00	−0.02	−0.07	−0.01	−0.01
EOM/S	−0.25	0.16	0.12	−0.29	0.11	0.13
TDDFT/L	0.56	0.09	0.04	0.51	0.04	−0.04
TDDFT/M	0.53	0.11	0.03	0.45	0.04	−0.04
TDDFT/S	0.38	0.28	0.15	0.28	0.17	0.11
TDHF/L	−0.41	−0.07	−0.09	−0.46	−0.12	−0.07
TDHF/M	−0.46	−0.05	−0.11	−0.53	−0.11	−0.07
TDHF/S	−0.61	0.13	0.12	−0.72	0.01	0.06
CIS/L	−0.48	−0.08	0.00	−0.53	−0.12	−0.07
CIS/M	−0.53	−0.07	0.01	−0.59	−0.11	−0.07
CIS/S	−0.68	0.12	0.12	−0.78	0.01	0.06

TABLE VII. Target values, $\Delta E_{\text{real}}^{\text{low}}$, and ONIOM $\pi \rightarrow \pi^*$ transition energy errors (eV) for the two isomers of the substituted cyclopropene in Fig. 4. The numbers in parentheses are the $\Delta E_{\text{model}}^{\text{high}}$ errors.

	1-trifluoro-methyl-cyclopropene			3-trifluoro-methyl-cyclopropene		
	$\Delta E_{\text{real}}^{\text{low}}$	ONIOM-1	ONIOM-2	$\Delta E_{\text{real}}^{\text{low}}$	ONIOM-1	ONIOM-2
Target	7.09	(0.00)	(0.09)	7.48	(0.44)	(0.49)
EOM/M	−0.05	0.00	−0.03	−0.09	−0.04	−0.03
EOM/S	−0.46	0.30	0.31	−0.55	0.19	0.20
TDDFT/L	0.90	0.07	0.11	0.85	0.01	0.06
TDDFT/M	0.84	0.07	0.09	0.80	0.01	0.05
TDDFT/S	0.51	0.33	0.34	0.48	0.27	0.27
TDHF/L	0.42	−0.09	0.06	0.50	0.02	0.11
TDHF/M	0.36	−0.08	0.05	0.43	0.03	0.11
TDHF/S	0.08	0.16	0.26	0.10	0.23	0.27
CIS/L	0.11	−0.13	−0.01	0.19	−0.01	0.07
CIS/M	0.05	−0.13	−0.02	0.13	0.00	0.07
CIS/S	−0.29	0.16	0.24	−0.26	0.23	0.26

TABLE VIII. Target values, $\Delta E_{\text{real}}^{\text{low}}$, and ONIOM errors (eV) for the two excited states of hexafluoro-1,3-butadiene in Fig. 5. The numbers in parentheses are the $\Delta E_{\text{model}}^{\text{high}}$ errors.

	A symmetry			B symmetry		
	$\Delta E_{\text{real}}^{\text{low}}$	ONIOM-1	ONIOM-2	$\Delta E_{\text{real}}^{\text{low}}$	ONIOM-1	ONIOM-2
Target	6.71	(−0.25)	(−0.11)	6.75	(−0.22)	(−0.05)
EOM/M	0.05	−0.02	0.06	0.00	−0.07	0.06
EOM/S	−0.34	0.49	0.51	−0.18	0.64	0.47
TDDFT/L	1.00	0.35	0.67	1.07	0.42	0.09
TDDFT/M	1.01	0.31	0.71	1.07	0.37	0.10
TDDFT/S	0.85	1.05	0.51	1.01	1.21	0.41
TDHF/L	−0.59	−0.26	−0.55	−0.25	0.08	−0.58
TDHF/M	−0.57	−0.25	−0.46	−0.26	0.05	−0.54
TDHF/S	−0.89	0.08	0.55	−0.36	0.60	−0.19
CIS/L	−0.66	−0.27	−0.58	−0.47	−0.09	−0.58
CIS/M	−0.63	−0.26	−0.48	−0.48	−0.11	−0.55
CIS/S	−0.97	0.09	0.64	−0.60	0.45	−0.12

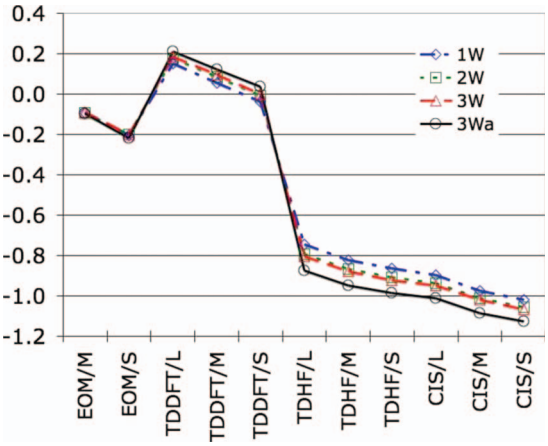


FIG. 6. Errors (eV) for the $\Delta E_{\text{real}}^{\text{low}}$ calculations on the water clusters in Fig. 1.

regions, of which the corresponding model systems do not directly interact. Tschumper's multicenter ONIOM method is a generalization of such an approach but has been applied only to ground state calculations.[50] Table V contains the target value (0.39 eV) and the errors. For the $\Delta E_{\text{real}}^{\text{low}}$ calculations the HF based methods perform all quite well, with errors $\leq$0.05 eV (4%–13%), unless the small basis set is used. The excited states are shifted, usually higher in energy, with respect to the target, but the gap between the states is well reproduced (compare also Table IV). In contrast, TDDFT shows deviations larger than 0.1 eV, which are significant considering the absolute $\Delta E_{1\to 2}^{\text{target}}$ ($\geq$30%). The $(\Delta E_{\text{model}}^{\text{high}})_{1\to 2}$ errors are large, between (-0.21)–(-0.44) eV, thus even larger than the target value (54%–114%). On the other hand, the ONIOM errors with the L and M basis sets are smaller (0%–34%). ONIOM considerably improves the $(\Delta E_{\text{model}}^{\text{high}})_{1\to 2}$-only approximation with all the methods. When the HF methods are used in the low level, the cancellation of errors of the conventional calculations is lost, but the ONIOM errors remain small. As for TDDFT, ONIOM leads to a considerable improvement over both $(\Delta E_{\text{model}}^{\text{high}})_{1\to 2}$ and $(\Delta E_{\text{low}}^{\text{real}})_{1\to 2}$. The relative errors are large for $\Delta E_{1\to 2}$ as the target value is small (0.39 eV). However, ONIOM behaves well compared to the other approximations. In particular the combination 4-1 (the smallest models for both the transitions) with TDDFT in the low level presents small errors with L and M basis sets, -0.07 and -0.05 eV (17% and 13%), respectively.

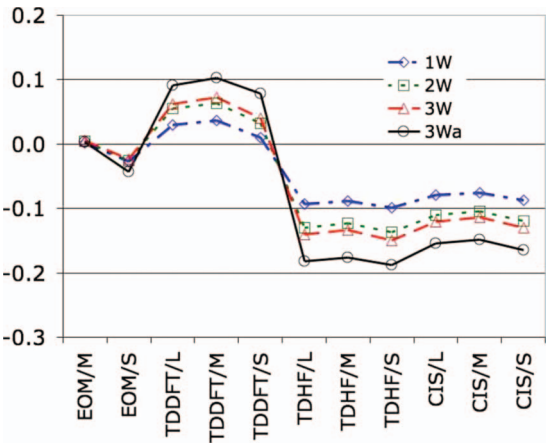


FIG. 7. Errors (eV) for the ONIOM calculations on the water clusters in Fig. 1.

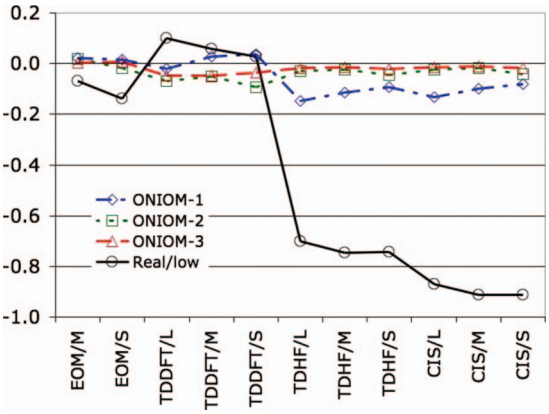


FIG. 8. Errors (eV) for *S*-azetidine-2-carboxylic acid in Fig. 2.

4. Trifluoro-methyl-cyclopropenes

In this section we analyze two similar molecules, cyclopropenes with a CF_3 group located on either position 1 or position 3 [see Figs. 4(a) and 4(d)]. We consider the first two electronic transitions, $\sigma\to\pi^*$ and $\pi\to\pi^*$, respectively, in both systems. Figures 4(b)–4(c) and 4(e)–4(f) show the models used in the ONIOM calculations. These molecules have an equivalent model system when the whole CF_3 group is eliminated (model 1), except for small differences in the geometry. In model 2, the cut is introduced on the C–F bonds, which is, again, not recommended in ONIOM calculations. Note that models 2 are no longer similar for the two molecules.

For the first transition, the $\Delta E_{\text{model}-1}^{\text{high}}$ errors in Table VI indicate that the substituent effect is very different in the two systems, namely, that the CF_3 group in position 1 does not affect the excitation as much as it does in position 3. $\Delta E_{\text{model}-2}^{\text{high}}$ errors show that model 2 alone does a very poor job in describing the real systems in both cases. This is not surprising since three polar C–F bonds are replaced by C–H. For the low cost methods, the absolute errors for the $\Delta E_{\text{real}}^{\text{low}}$ calculations are quite large for both molecules, 0.28–0.78 eV

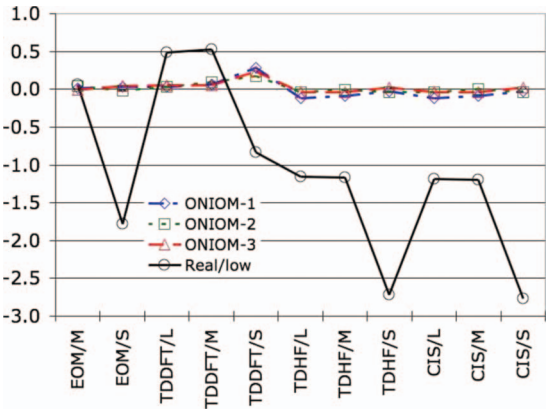


FIG. 9. Errors (eV) for the first transition of *R*-5-aminomethyl-2-oxazolidinone in Fig. 3.

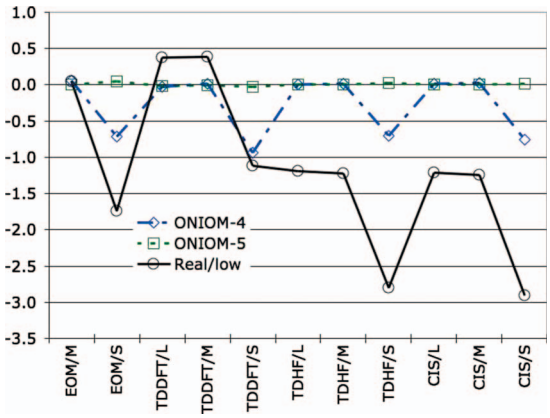


FIG. 10. Errors (eV) for the second transition of *R*-5-aminomethyl-2-oxazolidinone in Fig. 3.

in Table VI (4%–11%). The ONIOM-1 energies are considerably closer to the target than the corresponding $\Delta E^{\text{low}}_{\text{real}}$ for both molecules. Furthermore, the ONIOM-1 errors for both molecules are comparable and small (below 0.1 eV for TDDFT/L, 1%), despite the large difference in the $\Delta E^{\text{high}}_{\text{model}}$ errors (0%–4%). For ONIOM-2, the improvement is considerable for both molecules with respect to both the $\Delta E^{\text{high}}_{\text{model}}$ and the $\Delta E^{\text{low}}_{\text{real}}$ calculations, which indicates that the error due to the link atoms in the model systems indeed cancels between the high and low levels. With EOM-CCSD/M in the low level, ONIOM improves upon both $\Delta E^{\text{high}}_{\text{model}}$ and $\Delta E^{\text{low}}_{\text{real}}$ for both systems and both models, with errors $\leq$0.03 eV (see Table VI). The S basis set yields large $\Delta E^{\text{low}}_{\text{real}}$ errors, which reflect also in the ONIOM performance.

For the second transition ($\pi \rightarrow \pi^*$) the considerations outlined above are similar but amplified. Considering model 1, for the first molecule the ONIOM errors are small, but only equal to $\Delta E^{\text{high}}_{\text{model}}$ when EOM-CCSD/M is used as low level. For the second molecule $\Delta E^{\text{high}}_{\text{model}-1}$ is far from the target (more than 0.4 eV) and the ONIOM performances are excellent, $\leq$0.04 eV ($\leq$1%) with all the methods and the L and M basis sets. Also for this transition the ONIOM results seem to be more reliable than the $\Delta E^{\text{high}}_{\text{model}}$ or the $\Delta E^{\text{low}}_{\text{real}}$ alone, as the ONIOM errors stay small independently on how close (or far) the stand alone calculations are from the target. When model 2 is used in the first molecule, the ONIOM errors are smaller than for $\Delta E^{\text{high}}_{\text{model}}$ when HF based methods and the larger basis sets are used. In the second molecule ONIOM performs considerably better than $\Delta E^{\text{high}}_{\text{model}}$ and $\Delta E^{\text{low}}_{\text{real}}$, with TDDFT errors of $\leq$0.06 eV.

TABLE IX. Error rms and MAE for all the molecular systems. The error is defined in Eq. (3). The ONIOM values only include the models that are the most balanced choice for each molecular system (when more than one model is defined). These models are discussed in the text, and they are signaled with an (*) in Figs. 2–4. The data set includes the difference in the transition energies between the two excited states of *R*-5-aminomethyl-2-oxazolidinone, $\Delta E_{1\rightarrow 2}=\Delta E_2-\Delta E_1$ (Sec. III B 3). The contribution from hexafluoro-1,3-butadiene is not included (see text for details).

	rms		MAE	
	$\Delta E^{\text{low}}_{\text{real}}$	ONIOM	$\Delta E^{\text{low}}_{\text{real}}$	ONIOM
$\Delta E^{\text{high}}_{\text{model}}$		0.34		0.61
EOM/M	0.13	0.02	0.22	0.06
EOM/S	0.76	0.32	1.78	0.75
TDDFT/L	0.47	0.06	0.90	0.09
TDDFT/M	0.44	0.06	0.84	0.11
TDDFT/S	0.48	0.48	1.12	1.22
TDHF/L	0.74	0.11	1.19	0.18
TDHF/M	0.80	0.10	1.22	0.19
TDHF/S	1.23	0.31	2.80	0.71
CIS/L	0.81	0.11	1.21	0.15
CIS/M	0.87	0.10	1.24	0.16
CIS/S	1.31	0.33	2.91	0.76

5. Hexafluoro-1,3-butadiene

The final system we study is hexafluoro-1,3-butadiene.[51–54] It is not planar and belongs to the point group C_2, as shown in Fig. 5(a). This yields two classes of excited states of symmetries A and B. We analyze the first excited state for each symmetry.

According to the ONIOM rules, all of this molecule should be included in the model system. Nevertheless, we attempt to define two models, reported in Figs. 5(b) and 5(c). Each model necessarily implies making more drastic approximations than are normally recommended. Model 1 partitions the molecule in half so that the information about the symmetry is lost and the first excited states of both symmetries become identical. In model 2 all the fluorine atoms, which have an active role in the delocalized electronic distribution, are replaced by hydrogens. This model may introduce a large link atom effect. Other models could be defined by eliminating selected fluorine atoms, but there is no intuitive or systematic way to make this decision. Therefore we did not explore this possibility.

For the A symmetry, the errors in the $\Delta E^{\text{low}}_{\text{real}}$ calculations are large, except for EOM-CCSD/M (Table VIII). TDDFT underestimates the target value by 1 eV (15%); CIS and TDHF overestimate the target by 0.6 eV (8%–10%). The $\Delta E^{\text{high}}_{\text{model}}$ errors are also quite large, −0.25 and −0.11 eV for models 1 and 2, 4% and 2%, respectively. As for the ONIOM errors, EOM-CCSD/M presents values smaller than 0.1 eV. However, the improvement from the $\Delta E^{\text{low}}_{\text{real}}$ approximation is small for model 1, and there is no improvement with model 2, and the $\Delta E^{\text{low}}_{\text{real}}$ calculation is computationally expensive. Also ONIOM-1 with CIS/S and TDHF/S as low levels provides total errors less than 0.1 eV, probably due to cancellation of errors. All the other ONIOM combinations produce very large errors.

As for the B symmetry, the EOM-CCSD/M $\Delta E^{\text{low}}_{\text{real}}$ is extremely close to the target. EOM-CCSD/S is quite close, too. TDHF/L and M are not too far from the target value, whereas CIS is further (4%–7%). TDDFT performs very unsatisfactorily (16%) with all the basis sets (Table VIII). The $\Delta E^{\text{high}}_{\text{model}}$ errors show that model 2 is a very good approximation for this transition. Model 1 is not so different than for the A symmetry transition because the symmetry produces the splitting of the first excited state in two states that are close in energy. ONIOM EOM-CCSD/M errors are small, but they are significantly larger than the $\Delta E^{\text{low}}_{\text{real}}$ and for model 2 even larger than the $\Delta E^{\text{high}}_{\text{model}}$. TDDFT results are quite good

for model 2 if compared with the $\Delta E^{\text{low}}_{\text{real}}$ but not with $\Delta E^{\text{high}}_{\text{model}}$. CIS and TDHF perform well with model 1 and very poorly with model 2.

There is no trend, as methods that perform well with one model do not do so with the other. In addition, different methods do not perform consistently well with the same model. Therefore, for this system, there is no combination of model/method/basis set that works comparably well for both the transitions considered. Thus, for this case, the usual rules for the selection of a model system are not only sufficient but also necessary.

C. General trends

We now discuss the results and trends that are common to all the test cases. We separate the analysis of EOM-CCSD as low level from the other methods, TDDFT, CIS, and TDHF, since these are much less expensive methods (low cost) than the former (high cost). Thus, they are more likely used in production calculations, where the real system might be far larger than the model system and therefore not treatable at EOM-CCSD level with any basis set. However, cases may arise when the same level of theory is used at the high and low level, for instance, when the target basis set is too large for the real system, but a smaller one can still be manageable. In such cases, a precise approximation of the target calculation can be achieved with ONIOM.

The ONIOM rms/MAE errors with the low cost methods are always much smaller than the $\Delta E^{\text{low}}_{\text{real}}$ ones, except for TDDFT/S. The ONIOM extrapolation using the medium and large basis sets with all the low level methods significantly improves upon just the $\Delta E^{\text{high}}_{\text{model}}$ (error rms of 0.34 eV). TDDFT at low level (with M and L basis sets), in conjunction with EOM-CCSD on the model system, performs extremely well, with an rms error of 0.06 eV compared to more than 0.4 eV for $\Delta E^{\text{low}}_{\text{real}}$. Even CIS and TDHF provide good results when used as the low level, showing a deviation from the target around 0.1 eV. When CIS, TDHF, and TDDFT are used as the low level in ONIOM, the total time is dominated by the $\Delta E^{\text{high}}_{\text{model}}$ term.

There is no particular reason to prefer a basis set smaller than the large (6-311+G(d,p)) in the low level, as the limiting step in the overall ONIOM calculation is the $\Delta E^{\text{high}}_{\text{model}}$. The difference between the large and medium basis sets at the same level of theory is in general smaller than the effect of the level of theory itself in the ONIOM calculations (Table IX). On the other hand, the small basis set without diffuse functions is often not sufficiently accurate to describe electronic transitions with any method. Although in some cases very small errors are found, as in Sec. III B 1 for TDDFT (Table II) or in Sec. III B 3 for CIS and TDHF (first transition, Table IV), they are the results of error cancellation with the low level approximation.

We examine different partitioning for each example. In some cases chemical bonds involved in a ring and polar (i.e., non-C–C) bonds are cut. These cases may represent constrained bonds and are less well approximated by a hydrogen link atom. Although we do not necessarily recommend such partitions for production calculations, they do allow us to investigate the performance of ONIOM at the limits of its applicability. For all the sections, besides Sec. III B 1 where the high and the low regions are not covalently bonded, the smallest models are often the preferred choices since they yield errors that are not much larger than with larger models but provide considerable CPU time savings compared to the calculations with larger models. They always include the fewest number of centers directly involved in the transition under consideration and, except model 4 in Sec. III B 3, only involve cutting C–C bonds. In the case of Sec. III B 3, model 4 is obtained cutting the five membered ring, but it is preferred due to the saving in computational time. Furthermore, in this case the effect of the link atom is small, as shown by the small error in the $\Delta E^{\text{high}}_{\text{model}}$ term (−0.05 eV, Table IV). When $\Delta E^{\text{high}}_{\text{model}}$ is close to the target and hence the substituent effect is small, ONIOM does not always manage to improve the results, as discussed in Sec. III B 4. This is because the errors intrinsic in the extrapolation scheme are of the same order of magnitude of the $\Delta E^{\text{high}}_{\text{model}}$ error. Thus the results are not worsened and the overall error remains small.

An interesting aspect of the choice of the model and of the comparison of ONIOM to conventional calculations is analyzed in Sec. III B 4, where the two molecules differ only in the position of the CF_3 group on the central ring structure. The same model (model 1) can be used for the two molecules, albeit with small geometrical differences. When treated conventionally at the high level of theory, this model turns out to be a very good approximation of the target calculation on the first molecule but much worse on the second one ($\Delta E^{\text{high}}_{\text{model}}$ errors for the first transition are −0.03 and 0.32 eV, respectively, Table VI). In contrast, when ONIOM is used, the errors are small for both cases (below 0.1 eV for TDDFT/L as low levels, Table VI), showing that the ONIOM results do not necessarily depend on the accuracy of either the conventional $\Delta E^{\text{high}}_{\text{model}}$ or $\Delta E^{\text{low}}_{\text{real}}$ calculation. This also applies to the ONIOM results obtained with CIS and TDHF in the low level, as shown in Table IX, where, for instance, the rms error for CIS/L goes from 0.81 eV for $\Delta E^{\text{low}}_{\text{real}}$ to 0.11 eV with ONIOM, and the MAE goes from 1.21 to 0.15 eV.

EOM-CCSD/M is the best low level method (0.02 eV error rms). However, it also requires a CPU expensive calculation for the $\Delta E^{\text{low}}_{\text{real}}$ term, which often is the bottleneck calculation, especially when small model systems are considered (in contrast to the low cost methods). Thus, if the CPU time is not the limiting factor that prevents a full calculation at the target level, but other hardware factors are, as disk or memory, then ONIOM with EOM-CCSD/M as low level may allow a very accurate estimate of the target result. Although EOM-CCSD/S as low level can be considered a good alternative to EOM-CCSD/M from the CPU time point of view, its rms is large (0.32 eV) since even with this level of theory, the diffuse functions are sometimes necessary even for a qualitative description of the transition.

Furthermore, we note that it is important to pay much care in matching the excited states between the different models/methods/basis sets. In fact, different methods and different basis sets (in different molecules) can yield a different excited state order. An analysis of this kind by hand is relatively easy for the first few excited states, but for higher

excited states it becomes increasingly important to automate this, for instance, by comparing changes in the electron density populations upon excitation to each state. However, this is not a severe limitation in many practical cases, as one is often interested in a few low lying excited states.

Finally, Sec. III B 5 illustrates an example where the ONIOM approach fails seriously. It is the case where all the atoms are involved in the transition process and the definition of a sensible model system implies that the entire molecule must be in the model system. Various attempts are made to define model systems, but none improves the accuracy because the models are too approximate. This case clearly shows the limits of the range of applicability of a hybrid method such as ONIOM.

IV. SUMMARY AND CONCLUSIONS

In this work a series of benchmarking calculations is performed in order to test the ability of the ONIOM hybrid method to reproduce vertical electronic transition energies at a high level of theory, EOM-CCSD/6-311+G(d,p). The test molecules are chosen such that they include a large range of features that may affect the ONIOM results.

Based on the tests considered in this paper, we conclude that ONIOM is able to accurately approximate the transition energy computed at a high and computationally expensive level of theory such as EOM-CCSD, provided that the model system includes the atoms primarily involved in the excitation. In this way it is possible to define a model system that can be treated at the high level of theory, while the remaining substituent effects can be treated at a lower and less resource-demanding level. The model can be defined in various ways, but there is no particular improvement in the accuracy of the results by increasing the size of the model beyond the group mainly responsible for the transition. In general, the same rules for choosing the ONIOM model system for ground state problems are found to apply to electronic excitation as well.

The total CPU time is primarily determined by the $\Delta E_{\text{model}}^{\text{high}}$ calculation when methods such as CIS, TDHF, or TDDFT are used as low level. Among them, TDDFT performs best and can be considered the preferred choice as low level method, as it provides a very good balance between accuracy and computational effort.

[1] H. Sekino and R. J. Bartlett, Int. J. Quantum Chem. **18**, 255 (1984).
[2] J. Geertsen, M. Rittby, and R. J. Bartlett, Chem. Phys. Lett. **164**, 57 (1989).
[3] J. F. Stanton and R. J. Bartlett, J. Chem. Phys. **98**, 7029 (1993).
[4] R. J. Bartlett and M. Musial, Rev. Mod. Phys. **79**, 291 (2007).
[5] S. R. Gwaltney and R. J. Bartlett, Chem. Phys. Lett. **241**, 26 (1995).
[6] S. R. Gwaltney, M. Nooijen, and R. J. Bartlett, Chem. Phys. Lett. **248**, 189 (1996).
[7] M. Kallay and J. Gauss, J. Chem. Phys. **121**, 9257 (2004).
[8] H. J. Monkhorst, Int. J. Quantum Chem. **Y11**, 421 (1977).
[9] H. Koch and P. Jorgensen, J. Chem. Phys. **93**, 3333 (1990).
[10] S. Dapprich, I. Komaromi, K. S. Byun, K. Morokuma, and M. J. Frisch, J. Mol. Struct.: THEOCHEM **461**, 1 (1999).
[11] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, J. Phys. Chem. **100**, 19357 (1996).
[12] S. Humbel, S. Sieber, and K. Morokuma, J. Chem. Phys. **105**, 1959 (1996).
[13] T. Vreven and K. Morokuma, J. Comput. Chem. **21**, 1419 (2000).
[14] T. Vreven and K. Morokuma, *Annual Reports in Computational Chemistry* (Elsevier, Oxford, 2006), Vol. 2, Chap. 3, pp. 37–53.
[15] K. Morokuma, D. G. Musaev, T. Vreven, H. Basch, M. Torrent, and D. V. Khoroshun, IBM J. Res. Dev. **45**, 367 (2001).
[16] T. Vreven, K. Morokuma, O. Farkas, H. B. Schlegel, and M. J. Frisch, J. Comput. Chem. **24**, 760 (2003).
[17] T. Vreven, K. S. Byun, I. Komaromi, S. Dapprich, J. A. Montgomery, K. Morokuma, and M. J. Frisch, J. Chem. Theory Comput. **2**, 815 (2006).
[18] M. J. Bearpark, F. Ogliaro, T. Vreven, M. Boggio-Pasqua, M. J. Frisch, S. M. Larkin, M. Morrison, and M. A. Robb, J. Photochem. Photobiol., A **190**, 207 (2007).
[19] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).
[20] U. C. Singh and P. A. Kollman, J. Comput. Chem. **7**, 718 (1986).
[21] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[22] T. Vreven and K. Morokuma, J. Chem. Phys. **113**, 2969 (2000).
[23] S. P. Walch, Chem. Phys. Lett. **374**, 501 (2003).
[24] R. B. Zhang, X. C. Ai, X. K. Zhang, and Q. Y. Zhang, J. Mol. Struct.: THEOCHEM **680**, 21 (2004).
[25] R. Casadesus, M. Moreno, and J. M. Lluch, J. Photochem. Photobiol., A **173**, 365 (2005).
[26] C. Raynaud, R. Poteau, L. Maron, and F. Jolibois, J. Mol. Struct.: THEOCHEM **771**, 43 (2006).
[27] J. A. Gascon and V. S. Batista, Biophys. J. **87**, 2931 (2004).
[28] A. Yamada, T. Ishikura, and T. Yamato, Proteins: Struct., Funct., Bioinf. **55**, 1063 (2004).
[29] F. Blomgren and S. Larsson, J. Phys. Chem. B **109**, 9104 (2005).
[30] T. Vreven and K. Morokuma, Theor. Chem. Acc. **109**, 125 (2003).
[31] T. Vreven and K. Morokuma, *Continuum Solvation Models in Chemical Physics: From Theory to Application* (Wiley, Germany, 2007), Chap. 4.2, pp. 523–537.
[32] D. M. J. Bearpark, S. M. Larkin, and T. Vreven, J. Phys. Chem. A **112**, 7286 (2008).
[33] T. Vreven, M. J. Frisch, K. N. Kudin, H. B. Schlegel, and K. Morokuma, Mol. Phys. **104**, 701 (2006).
[34] K. F. Hall, T. Vreven, M. J. Frisch, and M. J. Bearpark, J. Mol. Biol. **383**, 106 (2008).
[35] E. Derat, J. Bouquant, and S. Humbel, J. Mol. Struct.: THEOCHEM **632**, 61 (2003).
[36] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).
[37] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[38] P. J. Stephens, F. J. Devlin, C. S. Ashvar, C. F. Chabalowski, and M. J. Frisch, Faraday Discuss. **99**, 103 (1994).
[39] M. J. Frisch, G. W. Trucks, H. B. Schlegel, *et al.*, GAUSSIAN DEVELOPMENT VERSION, Revision G.01+, Gaussian, Inc., Wallingford, CT, 2008.
[40] See EPAPS supplementary material at http://dx.doi.org/10.1063/1.3236938 for the geometries of the real and model systems, the transition energy values for the target, real system, model system, and the ONIOM calculations, and the unsigned relative error.
[41] F. Couty, G. Evano, M. Vargas-Sanchez, and G. Bouzas, J. Org. Chem. **70**, 9028 (2005).
[42] D. Kern, M. Schutkowski, and T. Drakenberg, J. Am. Chem. Soc. **119**, 8403 (1997).
[43] Y. Futamura, M. Kurokawa, R. Obata, S. Nishiyama, and T. Sugai, Biosci., Biotechnol., Biochem. **69**, 1892 (2005).
[44] Z. Sajjadi and W. D. Lubell, J. Pept. Res. **65**, 298 (2005).
[45] E. Barak, S. Amin-Spector, E. Gerliak, S. Goyard, N. Holland, and D. Zilberstein, Mol. Biochem. Parasitol. **141**, 99 (2005).
[46] R. M. Moision and P. B. Armentrout, J. Phys. Chem. A **110**, 3933 (2006).
[47] B. J. Ludwig, W. A. West, and D. W. Farnsworth, J. Am. Chem. Soc. **76**, 2891 (1954).
[48] U. Singh, B. Raju, S. Lam, J. Zhou, R. Gadwood, C. Ford, G. Zurenko, R. Schaadt, S. Morin, W. Adams, J. M. Friis, M. Courtney, J. Palandra, C. J. Hackbarth, S. Lopez, C. Wu, K. H. Mortell, J. Trias, Z. Yuan, D. V. Patel, and M. F. Gordeev, Bioorg. Med. Chem. Lett. **13**, 4209 (2003).
[49] S. W. Kim, J. G. Lee, E. J. Lee, H. Y. P. Choo, C. Y. Yoo, D. Y. Lee, K. R. Roh, and E. K. Kim, J. Comb. Chem. **6**, 851 (2004).
[50] B. W. Hopkins and G. S. Tschumper, J. Comput. Chem. **24**, 1563 (2003).

[51] R. H. Pottier, G. P. Semeluk, and R. D. S. Stevens, Spectrosc. Lett. **2**, 369 (1969).
[52] M. S. C. Foley, D. A. Braden, B. S. Hudson, and M. Z. Zgierski, J. Phys. Chem. A **101**, 1455 (1997).
[53] H. G. Cho, K. W. Kim, and B. S. Cheong, Bull. Korean Chem. Soc. **25**, 452 (2004).
[54] H.-Y. Xiao, J. Cao, Y.-J. Liu, W.-H. Fang, H. Tachikawa, and M. Shiotani, J. Phys. Chem. A **111**, 5192 (2007).